

Using Response Models to Estimate Channel Capacity for Neuronal Classification of Stationary Visual Stimuli Using Temporal Coding

MATTHEW C. WIENER AND BARRY J. RICHMOND

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892-4415

Wiener, Matthew C. and Barry J. Richmond. Using response models to estimate channel capacity for neuronal classification of stationary visual stimuli using temporal coding. *J. Neurophysiol.* 82: 2861–2875, 1999. Both spike count and temporal modulation are known to carry information about which of a set of stimuli elicited a response; but how much information temporal modulation adds remains a subject of debate. This question usually is addressed by examining the results of a particular experiment that depend on the specific stimuli used. Developing a response model allows us to ask how much more information is carried by the best use of response strength and temporal modulation together (that is, the channel capacity using a code incorporating both) than by the best use of spike count alone (the channel capacity using the spike count code). This replaces dependence on a particular data set with dependence on the accuracy of the model. The model is constructed by finding statistical rules obeyed by all the observed responses and assuming that responses to stimuli not presented in our experiments obey the same rules. We assume that all responses within the observed dynamic range, even if not elicited by a stimulus in our experiment, could be elicited by some stimulus. The model used here is based on principal component analysis and includes both response strength and a coarse (± 10 ms) representation of temporal modulation. Temporal modulation at finer time scales carries little information about the identity of stationary visual stimuli (although it may carry information about stimulus motion or change), and we present evidence that, given its variability, it should not be expected to do so. The model makes use of a linear relation between the logarithms of mean and variance of responses, similar to the widely seen relation between mean and variance of spike count. Responses are modeled using truncated Gaussian distributions. The amount of stimulus-related information carried by spike count in our data are 0.35 and 0.31 bits in primary visual and inferior temporal cortices, respectively, rising to 0.52 and 0.37 bits for the two-principal-component code. The response model estimates that the channel capacity is 1.1 and 1.4 bits, respectively, using the spike count only, rising to 2.0 and 2.2 bits using two principal components. Thus using this representation of temporal modulation is nearly equivalent to adding a second independent cell using the spike count code. This is much more than estimated using transmitted information but far less than would be expected if all degrees of freedom provided by the individual spike times carried independent information.

INTRODUCTION

It is not yet completely understood how information is encoded in neuronal spike trains and how much information is carried. In the visual system, it is clear that the number of

action potentials elicited by a visual stimulus is an important part of the code for carrying stimulus-related information. There is now strong evidence that modulation of the firing rate during the course of the response to a stimulus presented for the period of a typical intersaccadic interval (~ 300 ms) carries additional information that is not available from average response strength alone (Heller et al. 1995; Richmond and Optican 1987, 1990; Tovee et al. 1993; Victor and Purpura 1996). Here we investigate how much information temporal modulation adds.

As often understood, the answer to this question depends on the experiment from which we take our data. Different sets of stimuli will elicit different sets of responses, which may encode more or less information using temporal modulation. Using response models can help us answer a question less tied to a particular experiment: how much more information can be carried by the best use of temporal modulation than by the best use of spike count alone? The maximum amount of information that can be carried by a neuron using a particular code (such as spike count, or spike count along with some representation of temporal modulation) is called its channel capacity (Cover and Thomas 1991; Shannon and Weaver 1949). The channel capacity is defined uniquely for a particular neuron using a particular code. Channel capacity's uniqueness comes at a price, however: to calculate channel capacity, we must know all possible responses that can be elicited from a neuron rather than only those observed in our experiment.

Our goal is to construct a response model that not only describes the responses observed in an experiment but also can be used to predict the responses that would be elicited by other stimuli. This substitutes dependence on the accuracy of the model for the limitations imposed by the amount of data typically collected in experiments. Ideally we would like this model to include all features of the response that do carry unique information, without complicating matters by including features that carry little or no unique information. Recently Gershon et al. (1998) presented an approach to constructing such a response model involving spike count only. The model used a widely known relation between mean and variance of spike count (Dean 1981; Lee et al. 1998; O'Keefe et al. 1997; Tolhurst et al. 1981, 1983; van Kan et al. 1985; Vogels et al. 1989), and the fact that distributions of spike count were well fit by a truncated Gaussian. This model allowed Gershon et al. (1998), given any mean spike count, to describe the set of responses giving rise to that mean. The set of responses corresponding to a given mean did not depend on whether the mean had actually been elicited by a stimulus in the experi-

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

ment. Thus the model predicted the responses associated with any possible mean. Because any set of responses has a mean, the model described all sets of responses that could be elicited from the neuron. This allowed Gershon et al. to estimate the neuron's channel capacity using the spike count code.

Here we extend the approach of Gershon et al. (1998) to a code that includes the temporal patterns of rate modulation of the responses. We represent the neuronal responses using the first two principal components. The coefficient with respect to the first principal component is strongly correlated with spike count, and the coefficient with respect to the second principal component is a coarse (± 10 ms) measure of the time course of a response, often indicating whether spikes tend to be concentrated at the beginning of a response (Richmond and Optican 1987, 1990). We identify relations among the means and variances of the principal component coefficients similar to the relation between mean and variance of spike count. We also find that distributions of the principal component coefficients are well fit by truncated Gaussian distributions. Now a set of responses must be described by two means—one for each component of the code. As in Gershon et al. (1998), this allows us to predict both the response strength and the temporal patterns of rate modulation for all possible sets of responses.

We estimated channel capacity on the basis of the response models for spike count and for the first two principal components. We found that the two-component code on average increases the channel capacity by 0.8–0.9 bits over that carried by spike count alone (1.1 and 1.4 bits, respectively) in neurons in the primary visual cortex (V1) and area TE of inferior temporal cortex. Therefore the contribution temporal modulation can make to neural information processing for identification of stationary stimuli, although far smaller than if all degrees of freedom in a spike train actually were used to carry such information, is substantially larger than would be estimated from only the responses seen in our experiment. This contrasts with situations in which rapidly changing stimuli drive precisely time-locked neuronal responses, which can lead to extremely high information transmission rates (Buracas et al. 1998). Finally, we present evidence that very little additional channel capacity for identifying stationary stimuli can be expected by using more principal components (that is, by representing the rate variation with higher precision).

METHODS

Data sets

We performed new analyses using previously published data. The data came from two studies of supragranular V1 complex cells, each study using two rhesus monkeys performing a simple fixation task (Kjaer et al. 1997; Richmond et al. 1990), and from one study of neurons in IT cortex in two other monkeys performing a simple sequential nonmatch-to-sample task (Eskandar et al. 1992). In these three studies, the visual stimuli were two-dimensional black-and-white patterns based on the Walsh functions (Fig. 1).

In both V1 studies, stimuli were presented centered on the neuronal receptive fields, located in the lower contralateral visual field 1–3° from the fovea. Stimuli covered the excitatory receptive field. At 3° eccentricity, the stimuli were $\sim 2.5^\circ$ on a side. In the first V1 study (V1 set 1) (Richmond et al. 1990), 128 stimuli were used: sixty-four 8×8 pixel patterns and their contrast-reversed counterparts. For V1 set 2 (Kjaer et al. 1997), 16 stimuli were used: a set of eight 16×16 pixel patterns and their contrast-reversed counterparts.

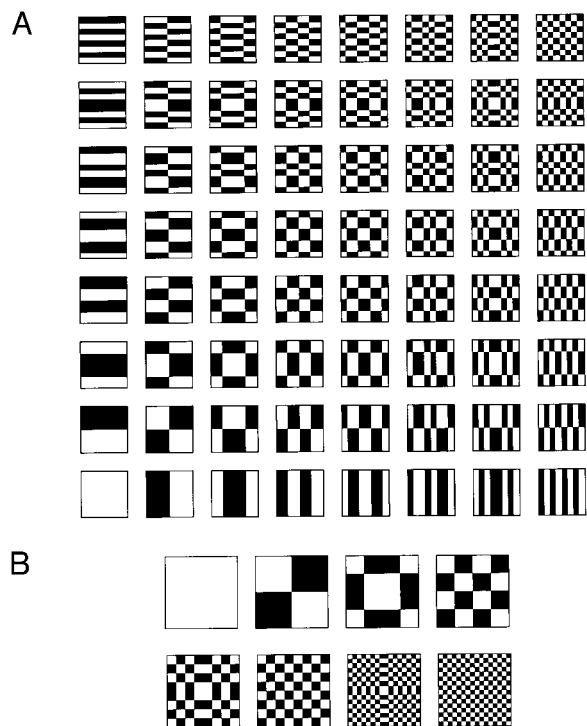


FIG. 1. Walsh patterns. For the primary visual cortex (V1) set 1, the 64 stimuli (A) and the corresponding contrast-reversed set were presented on the receptive fields while the monkeys fixated. For V1 set 2, the 8 stimuli (B) and the corresponding contrast-reversed set were presented on the receptive field while the monkey fixated. For both, the stimuli were 2.5° on a side (covering the excitatory receptive field and some of the surround). For the inferior temporal cortex (IT), the 4×4 set (16 stimuli) in the bottom left corner of A and the corresponding contrast-reversed set were used as the monkey performed a nonmatch-to-sample task. Stimuli were 4° on a side and were centered at the point of fixation.

In the IT study (Eskandar et al. 1992), stimuli were presented centered on the fovea. The patterns were 4° square. Thirty-two stimuli were used: sixteen 4×4 pixel patterns and their contrast-reversed counterparts.

The stimulus was displayed for 320 ms in V1 and 352 ms in IT. To account for response latencies and to avoid contamination from off-responses, spikes were counted during the interval from 30 to 300 ms after stimulus onset for the V1 neurons and 50 to 350 ms after stimulus onset for the IT neurons. Gershon et al. (1998) found that both transmitted information and channel capacity based on the spike count code are stable with respect to small changes in these counting windows. For each neuron, each stimulus was presented approximately the same number of times (± 2) in randomized order. Different neurons received different numbers of presentations. The number of presentations of each stimulus was between 10 and 50 in V1 and between 19 and 50 in IT. Seven V1 neurons with < 10 trials per stimulus were omitted from these analyses to ensure stability of information estimates (Golomb et al. 1997; Wiener and Richmond 1998). The timing of events, including spikes, was recorded with 1-ms resolution.

Quantifying the responses

Each spike train was low-pass filtered by convolution with a Gaussian distribution with standard deviation of 5 ms and resampled at 1-ms resolution to create a spike density function (Fig. 2, A and B). Convolution and resampling in this way avoids a problem of binned histograms, which do not distinguish between spikes at the center of a bin and those near the edges (Richmond et al. 1987; Sanderson and Kobler 1976; Silverman 1986).

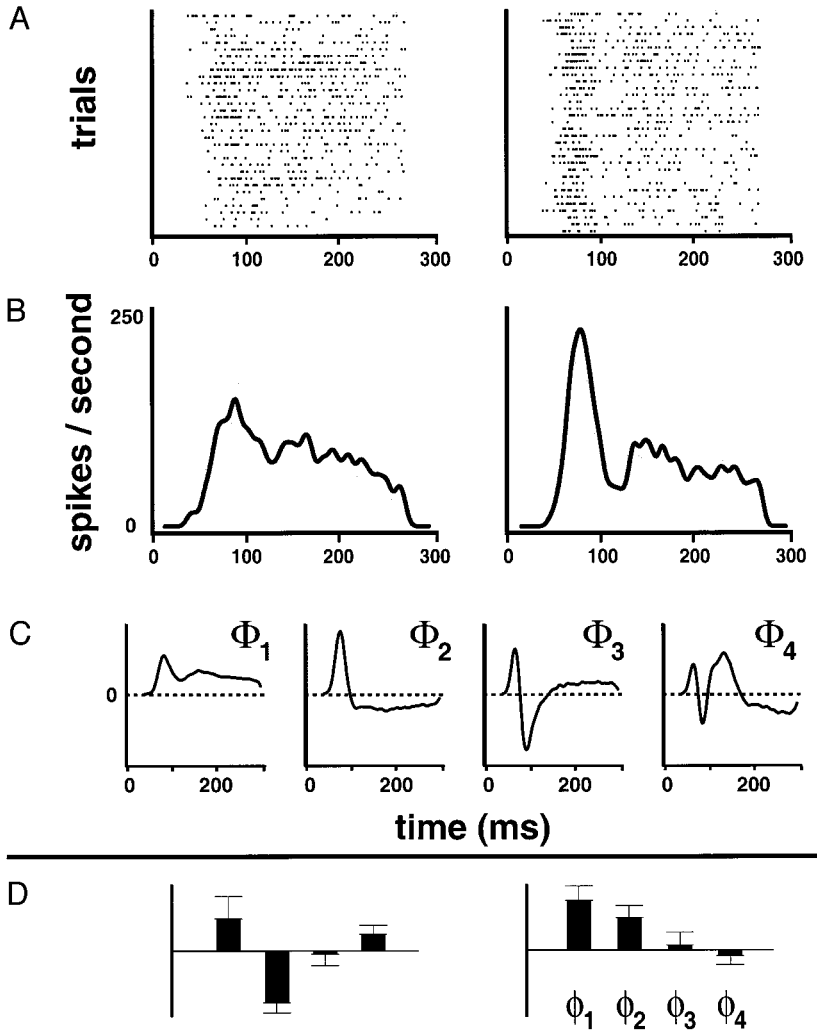


FIG. 2. Representation of spike trains using principal component analysis. *A*: rasters of the responses of a single V1 cell to 2 stimuli. Horizontal axis shows time from stimulus onset in milliseconds. Each tick mark represents an action potential; each row of tick marks represents a single stimulus-elicited response. *Left and right*: responses to different stimuli. Mean spike count is equal for the 2 sets of responses, but the responses to the 2nd stimulus (*right*) show a greater concentration of spikes occurring between ~ 50 and 120 ms after stimulus onset. *B*: spike density functions corresponding to the rasters above. Horizontal axis shows time from stimulus presentation (ms) on the same scale as for the rasters. Vertical axis shows the instantaneous firing rate at each time in spikes/s. Dark line shows mean firing rate at each time; the wider gray line shows one standard error above and below the mean. Concentration of spikes in the early part of the responses to the 2nd stimulus (*right*) is again apparent. *C*: 1st 4 principal components of the responses of the cell. Horizontal axis of each panel shows time, in milliseconds, after stimulus presentation. Note that a positive coefficient for the 2nd principal component indicates a high concentration of spikes near the beginning of a response. Scale of the principal components and the corresponding coefficients is arbitrary and therefore is omitted. *D*: principal component representation of the responses shown in the rasters and densities. Each bar plot shows the mean and standard deviation of the 1st 4 (*left to right*) principal component coefficients of the responses. Line connecting the bases of the bars shows 0. Note the difference in the 2nd principal component coefficients between the 2 panels. This difference reflects the fact that the spikes in the responses shown in the *right* are more concentrated at early times than are the spikes in the responses shown in the *left*. (In the calculations in the text the principal components are translated to always be positive; here we use principal components centered at 0 to emphasize the difference in the the coefficients with respect to the 2nd principal component.)

We used principal component analysis to reduce the dimension of the data set. Principal component analysis defines an ordered set of axes, each accounting for more of the variance in the data set than those that follow. These axes can be computed by finding the eigenvectors of the covariance matrix of the data (Ahmed and Rao 1975; Deco and Obradovic 1996). Each data point is defined uniquely by its projections onto these axes; these values are called the coefficients with respect to the principal components, or principal component coefficients. We use the code formed by the first k principal component coefficients because it is the optimal k -dimensional linear code for least-squares reconstruction and minimizes an upper bound of information loss through a one-layer network (Campa et al. 1995; Deco and Obradovic 1996; Plumbley 1991). Principal-component analysis has a long history in signal analysis (Ahmed and Rao 1975) and has been used to study information coded by temporal modulation in neuronal responses (Heller et al. 1995; Kjaer et al. 1994; McClurkin et al. 1991; Optican and Richmond 1987; Richmond and Optican 1987, 1990; Tovee and Rolls 1995; Tovee et al. 1993). Figure 2 shows rasters of the responses elicited by two stimuli from a single V1 neuron, the corresponding spike density functions, the first four principal components, Φ_1 – Φ_4 , of the responses from the neuron, and the principal component representation of the two responses.

In this study, each neuronal response was represented by the corresponding coefficients with respect to the first and second principal components of the spike density functions (Richmond and Optican 1987, 1990). We call these coefficients ϕ_1 and ϕ_2 . ϕ_1 and ϕ_2 can be translated by arbitrary constants as long as the appropriate constant

multiples of Φ_1 and Φ_2 are subtracted from the results. This manipulation is similar to rewriting the expression $a + bx$ as $a + bx_0 + b(x - x_0)$. It is conventional to use the average waveform as the base waveform, translating the axes so that the average waveform corresponds to a vector of zeros. This causes some values of ϕ_1 and ϕ_2 to be negative. Here we translated the axes so that ϕ_1 and ϕ_2 are always positive and logarithms could be taken. The specific form of the new base waveform is not important for our analyses.

Model of response variability

Estimating channel capacity requires estimating all possible (ϕ_1 , ϕ_2) response distributions. Our estimate of these distributions must be based on the responses elicited by stimuli actually presented in our experiments. It is important to note that if the responses elicited by the stimuli presented in our experiments are not representative of the responses elicited by other stimuli, our model will generalize poorly.

Ideally, we would estimate the two-dimensional conditional distribution of responses $p(\phi_1, \phi_2|s)$ directly for each stimulus s . However, we did not have enough responses to each stimulus to do so. Instead, we described the distributions of ϕ_1 and ϕ_2 individually, and assumed that ϕ_2 was independent of ϕ_1 (except for correlations imposed by truncation at the bounds of the response space, see following text). To be precise

$$p(\phi_1|s) = G(\mu_{1s}, \sigma_{1s}^2)/K_1(\phi_1) \quad \text{when} \quad \phi_1 \geq \phi_1^{\min} \\ = 0 \quad \text{otherwise} \quad (1)$$

and

$$p(\phi_2|s, \phi_1) = G(\mu_{2s}, \sigma_{2s}^2)/K_2 \quad \text{when} \quad \phi_2^{\min}(\phi_1) \leq \phi_2 \leq \phi_2^{\max}(\phi_1) \\ = 0 \quad \text{otherwise} \quad (2)$$

where G is the Gaussian distribution with indicated mean and variance, μ_{js} and σ_{js}^2 are the mean and variance of ϕ_j elicited by stimulus s , ϕ_1^{\min} is the minimum value of ϕ_1 , $\phi_2^{\min}(\phi_1)$ and $\phi_2^{\max}(\phi_1)$ are the calculated bounds on ϕ_2 for given ϕ_1 , and $K_1 = \int_{\phi_1^{\min}}^{\phi_1^{\max}} d\phi_1 p(\phi_1|s)$ and $K_2 = \int_{\phi_2^{\min}(\phi_1)}^{\phi_2^{\max}(\phi_1)} d\phi_2 p(\phi_2|s)$ are normalizing factors. We show in the results that the Gaussian distribution provides an acceptable fit to the distributions of ϕ_1 and ϕ_2 for each stimulus.

Although ϕ_1 and ϕ_2 are by construction uncorrelated in the responses from each neuron, correlations are sometimes present in the responses elicited by individual stimuli. In RESULTS, we observe that the set of responses is bounded; these bounds will be described quantitatively at that point. We observe correlations between ϕ_1 and ϕ_2 only in those distributions lying close to the response bounds. Therefore our model assumes that correlations between ϕ_1 and ϕ_2 arise only as a result of truncating distributions at the boundaries of the response space. We make this same assumption when characterizing the set of all possible response distributions to calculate channel capacity. We make this assumption because we do not have data to justify any other: under other circumstances (and with sufficient data) we might find, and include in a model, such correlations. The close match of our estimates of transmitted information to those obtained using other well-validated methods (see RESULTS) suggests that in this case the assumption is justified.

Means and variances of response distributions

Two-dimensional Gaussian distributions (given our correlation assumptions, preceding text) are characterized by two means and two variances. All four parameters can be measured from experimentally observed responses. In RESULTS, we show that the variances of ϕ_1 and ϕ_2 distributions are related to mean ϕ_1 by a power law

$$\sigma_i^2 = \exp(a_i) \mu_1^{b_i} \quad (3)$$

which is equivalent to

$$\log(\sigma_i^2) = a_i + b_i \log \mu_1 \quad (4)$$

where μ_i and σ_i^2 are the mean and variance of ϕ_i , for $i = 1, 2$. a_i and b_i are estimated by linear regression. These regressions were used to estimate σ_{is}^2 for $i = 1, 2$ in Eqs. 1 and 2. Note that the variances of both ϕ_1 and ϕ_2 were modeled as a function of the mean of ϕ_1 (we justify this in RESULTS).

Estimates of $\log(\mu)$ and $\log(\sigma^2)$ obtained by taking the logarithm of the sample mean and variance are biased, resulting in underestimation of the variance of response distributions and therefore overestimation of transmitted information. We corrected for the bias using a Taylor series expansion; only a few terms are needed for good results (Kendall and Stuart 1961, p. 4–6).

We also used linear regression to estimate the mean and variance of the values of ϕ_2 when ϕ_1 takes on a given value, no matter which stimulus elicited the response.

Transmitted information and channel capacity

The information carried in a neuron's response about which member of a set of stimuli is present (Cover and Thomas 1991) is defined as

$$I(R; S) = \sum_{s,r} p(s)p(r|s) \log_2 \frac{p(r|s)}{p(r)} \quad (5)$$

where S is the set of stimuli s , R is the set of responses r , $p(r|s)$ is the conditional probability of response r given stimulus s , $p(s)$ is the probability that stimulus s occurred, and $p(r) = \sum_s p(r|s)p(s)$ is the probability of response r . Equation 5 is general; it can be applied to responses r of any dimension. We used both the one-dimensional spike count code and the two-dimensional principal component code, $r = \phi = (\phi_1, \phi_2)$.

The transmitted information $I(\Phi; S)$ depends on $p(s)$, the distribution of presentation probabilities of the stimuli (see Eq. 5). The *channel capacity* of a cell is the maximum value of transmitted information over all distributions $p(s)$, where the set of stimuli S should now be understood to include all possible visual stimuli. Finding this maximum requires knowing the conditional response distributions $p(r|s)$ for all stimuli s . We estimate the distributions using the model described earlier.

From Eqs. 1 and 2, we see that estimating $p(\phi|\mu)$, the probability with which the distribution with mean $\mu = (\mu_1, \mu_2)$ elicits response $\phi = (\phi_1, \phi_2)$, requires μ_1 , μ_2 , σ_1^2 , and σ_2^2 . Given the power-law relations between the means and variances of distributions (Eq. 3), μ_1 determines both σ_1^2 and σ_2^2 but not μ_2 . Thus for each μ_1 there are many possible response distributions, identical except for translation in ϕ_2 . Each distribution, then, is characterized completely by the two means μ_1 and μ_2 . When calculating transmitted information, only the observed distributions are considered (Fig. 3, *top*); when calculating channel capacity, all possible distributions must be considered (Fig. 3, *bottom*).

According to this model, two stimuli that elicit the same μ_1 and μ_2 from a neuron in fact elicit identical response distributions and therefore cannot be distinguished from one another by that neuron. Thus a stimulus can be identified simply by the mean response it elicits. Equation 5 now can be written

$$I(\Phi; S) = \int d\mu p(\mu) \int d\phi p(\phi|\mu) \log_2 \frac{p(\phi|\mu)}{p(\phi)} \quad (6)$$

where $p(\mu)$ is the probability with which the distribution with mean μ is presented, $p(\phi|\mu)$ is the probability with which the distribution with mean $\mu = (\mu_1, \mu_2)$ elicits response $\phi = (\phi_1, \phi_2)$, and $p(\phi) = \int_\mu d\mu p(\phi|\mu)$ is the total probability with which response ϕ occurs. The channel capacity is the maximum of this expression over the two-dimensional distribution of probabilities $p(\mu) = p(\mu_1, \mu_2)$.

A derivation of Eq. 6 is given in Gershon et al. (1998). That derivation is expressed in terms of the spike count code but remains valid when spike count is replaced with any other response code. Information and channel capacity based on the spike count code were calculated using the method of Gershon et al. (1998).

The search for the maximizing set of probabilities is subject to three constraints: the probabilities must be nonnegative; the probabilities must sum to one; and the range of means must be finite. The first two constraints arise from intrinsic properties of probability distributions. If the third constraint is violated, the transmitted information can be infinite and the problem of maximizing transmitted information is ill-posed. The implementation of these constraints and other numerical issues are discussed in the APPENDIX.

We did not penalize distributions for probability weight falling outside the response envelope bounds for ϕ_2 (as we did for probability weight falling outside the observed range of ϕ_1); we simply ignored that portion of the distribution. This allows the widest possible separation between distributions, causing our estimates of channel capacity to be larger than if the boundaries had been strictly enforced. Thus our procedure was designed to give the most generous estimates possible (consistent with the constraints estimated from the data) of channel capacity associated with temporal coding.

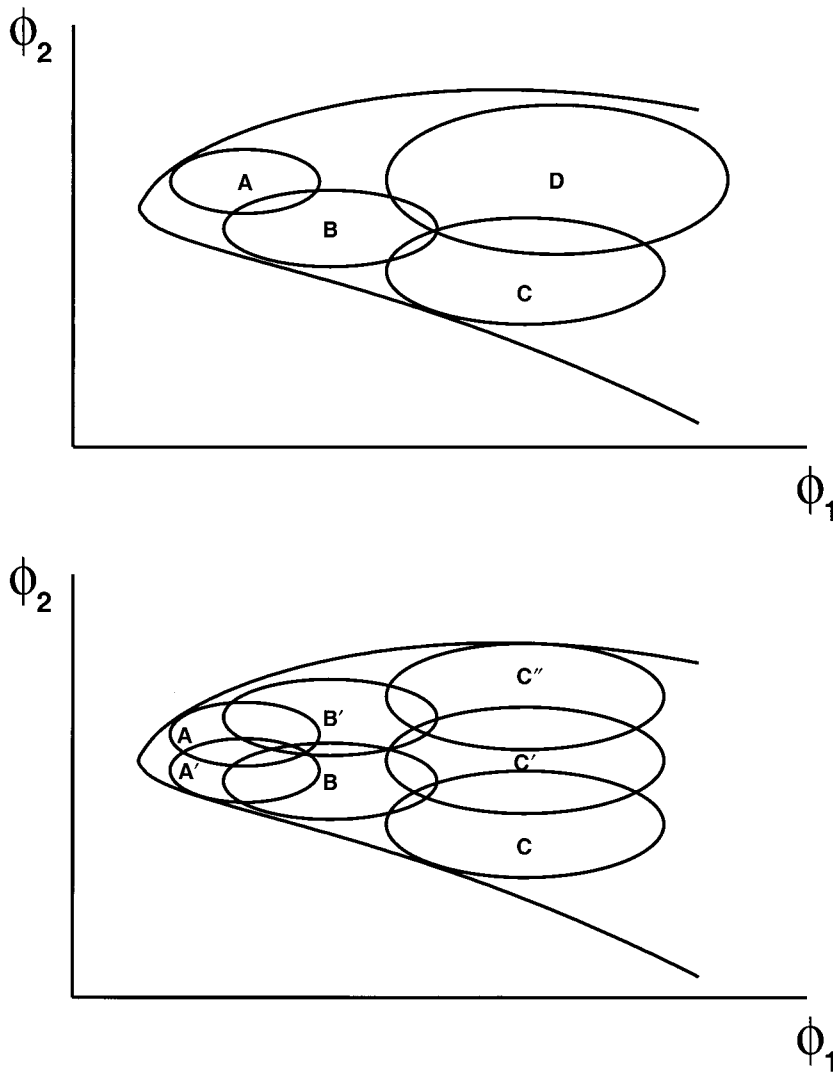


FIG. 3. Schematic of distributions used to calculate transmitted information and channel capacity. Lines represent the boundaries of the ϕ_1 - ϕ_2 envelope: region of ϕ_1 - ϕ_2 space in which responses fall. Ellipses represent distributions of responses elicited by various stimuli. *Top*: calculating transmitted information in an experiment. Response distributions elicited by 4 stimuli, A, B, C, and D, are shown. (These stimuli are merely illustrative and do not correspond to any stimuli from our actual experiments.) Mean values of ϕ_1 and ϕ_2 , μ_1 and μ_2 , of each distribution can be calculated from the data, and the variances of ϕ_1 and ϕ_2 , σ_1^2 and σ_2^2 , estimated from the mean of ϕ_1 using the regression relations. *Bottom*: when considering channel capacity, we must consider all possible pairs of means μ_1 , μ_2 . Value of μ_1 determines both σ_1^2 and σ_2^2 , so all distributions with a given value of μ_1 will have the same shape. Distributions elicited by A, B, and C (*top*) are shown, along with several other possible distributions with the same mean values of ϕ_1 . (Distribution elicited by stimulus D is omitted for clarity.)

RESULTS

We considered 6 neurons from one experiment in V1, 29 neurons from a second experiment in V1, and 19 neurons from IT. To calculate transmitted information and channel capacity, we characterized the space of responses, parameterized by ϕ_1 and ϕ_2 , the coefficients with respect to the first two principal components; determined the regression relations for the means and variances of ϕ_1 and ϕ_2 elicited by the stimuli; and constructed a model for the distributions of ϕ_1 and ϕ_2 .

ϕ_1 and ϕ_2 are somewhat abstract, but are related to biologically relevant aspects of the response. Responses with high (low) values of ϕ_1 correspond to responses with high (low) numbers of spikes (McClurkin et al. 1991; Richmond and Optican 1987, 1990; Tovee et al. 1993); in our data, the median correlation between spike count and ϕ_1 was 0.91 (interquartile range 0.86–0.97). ϕ_2 is a coarse (± 10 ms) measure of temporal modulation (Optican and Richmond 1987; Richmond and Optican 1987). It often characterizes whether or not spikes are concentrated in the early part of the response (Fig. 2).

Response space

Although the coefficients with respect to the first and second principal components (ϕ_1 and ϕ_2) are linearly uncorrelated by

construction (Ahmed and Rao 1975; Deco and Obradovic 1996), they are not independent. Figure 4 shows scatterplots of ϕ_2 versus ϕ_1 for two cells from V1 and two cells from IT. In these four cells, and in all other cells in our sample, the range of ϕ_2 increased with increasing ϕ_1 . The apex of the cone is the (ϕ_1 , ϕ_2) pair representing trials when no spikes were elicited.

The limited range of ϕ_2 for a given value of ϕ_1 reflects the fact that only a certain amount of temporal modulation is possible with a given number of spikes. If each of k spikes in a train could be placed in any of n time bins, assuming only that no bin could hold more than one spike, the number of possible patterns would be $\binom{n}{k} = n!/[k!(n-k)!]$. The range of values of ϕ_2 seen under this combinatorial assumption (Fig. 5, thin line) is much wider than the range observed experimentally for any given number of spikes k (Fig. 5, dots).

Because we do not understand the spike generation process well enough to predict the response envelope, we developed a statistical model of the mean and variance of distributions of ϕ_2 corresponding to different values of ϕ_1 . We divided the range of ϕ_1 into a number of bins, each of which defines a ϕ_1 slice in the two-dimensional response space. Figure 6A shows the *top left scatterplot* from Fig. 4 with slice boundaries superimposed. The mean and variance of ϕ_2 were calculated in

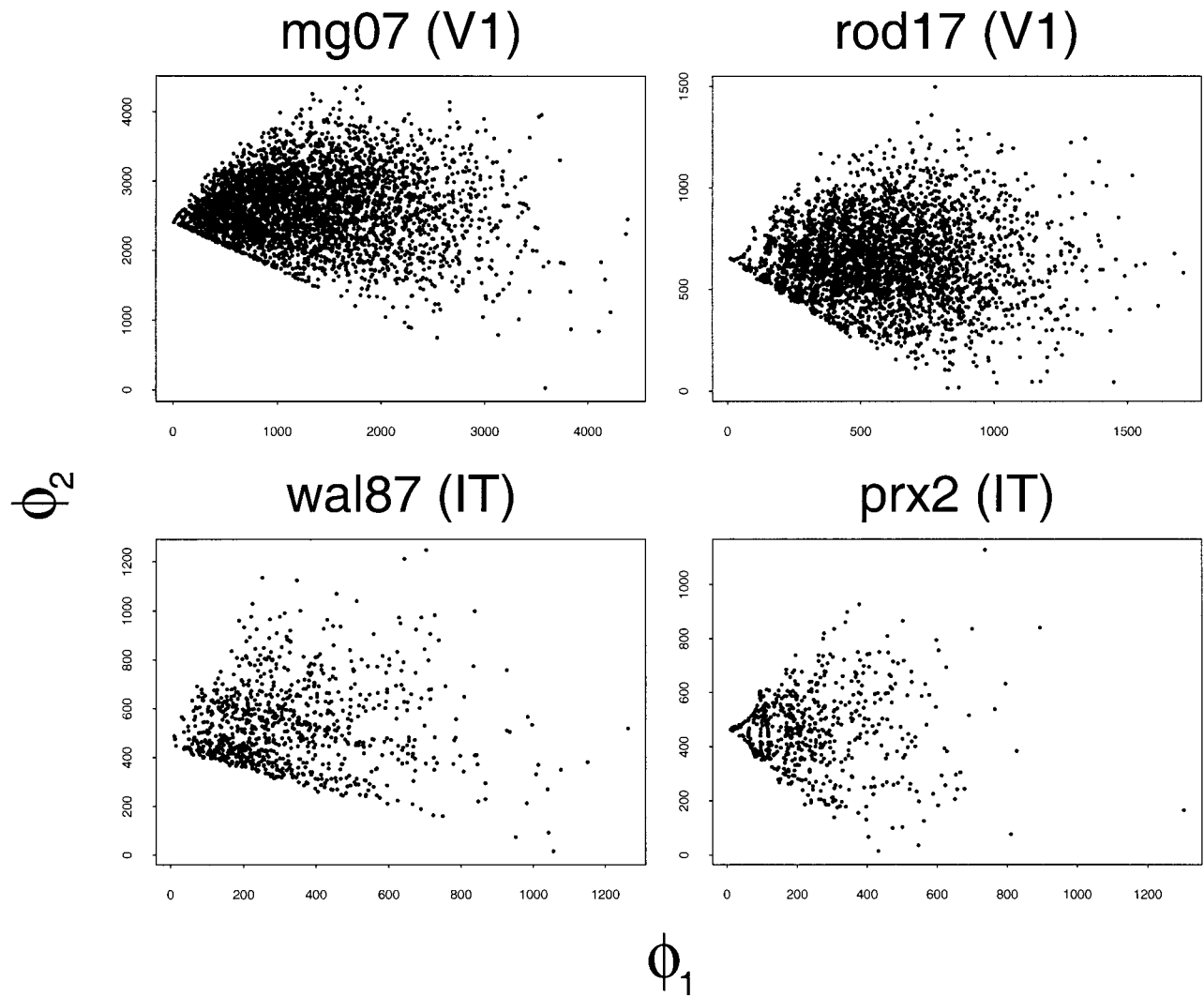


FIG. 4. Scatterplot of responses for 4 cells. For each panel, the horizontal axis shows ϕ_1 and the vertical axis shows ϕ_2 . Each point shows the response (ϕ_1, ϕ_2) for a single trial. Although ϕ_1 and ϕ_2 are linearly uncorrelated (by construction), they are not independent: range of ϕ_2 depends on the value of ϕ_1 .

each ϕ_1 slice. Figure 6B shows an example of the linear relation between $\log(\text{variance})$ of ϕ_2 and $\log(\text{mean})$ of ϕ_1 in each ϕ_1 slice. This relation was significant in 47/54 neurons at the $P < 0.01$ level (median $r^2 = 0.91$, $iqr = 0.79\text{--}0.94$). Figure 6C shows an example of a fit of the mean of ϕ_2 as a quadratic function of the mean of ϕ_1 in each slice. The quadratic fit was rejected in only 3/54 neurons (χ^2 test, $p \leq 0.01$) and is used here.

We used these relations to estimate the bounds of the envelope containing the response space by $\hat{\mu}_{\phi_2}(\phi_1) \pm k_{\pm} \hat{\sigma}_{\phi_2}(\phi_1)$, where $\hat{\mu}$ and $\hat{\sigma}$ are the mean and standard deviation estimated as in the preceding text. For each cell, k_+ (k_-) was chosen so that all but 0.1% of the points fell below (above) the boundary. Several points were allowed to fall outside the bounds to prevent excessive influence of outliers. We show below that small changes in the response envelope do not affect our results. For all 54 neurons, the response space envelope estimated using the regression (thick lines in Fig. 5) was much narrower than the envelope assuming spikes can fall in arbitrary 1-ms bins (the thin lines in Fig. 5).

Mean-variance relations of ϕ_1 and ϕ_2 by stimulus

The logarithms of the mean and variance of spike count in response to different stimuli are linearly related (Dean 1981; Gershon et al. 1998; Lee et al. 1998; O'Keefe et al. 1997; Tolhurst et al. 1981, 1983; van Kan et al. 1985; Vogel et al. 1989). Because ϕ_1 is correlated strongly with spike count (Richmond and Optican 1987), it is natural that the logarithms of the mean and variance of ϕ_1 also are related linearly (in 52/54 neurons at $P < 0.01$; median $r^2 = 0.79$, $iqr = 0.61\text{--}0.91$; see Fig. 7). By simple extension, we might expect there to be a linear relationship between $\log(\text{mean})$ and $\log(\text{variance})$ of ϕ_2 as well. However, such a relation existed in only 11/54 of the neurons ($P < 0.01$; median $r^2 = 0.09$, $iqr = 0.02\text{--}0.21$). Instead, we found that the variance of ϕ_2 elicited by a single stimulus increased with the mean of ϕ_1 (in 43/54 neurons at $P < 0.01$; median $r^2 = 0.66$, $iqr = 0.38\text{--}0.84$; see Fig. 8). This is consistent with the fact that the range (variance) of ϕ_2 increases with increasing ϕ_1 (Fig. 4). Adding $\text{mean}(\phi_2)$ to the model added very little explanatory power (median increase in $r^2 = 0.02$, $iqr = 0.0\text{--}0.09$); for simplicity, we omitted

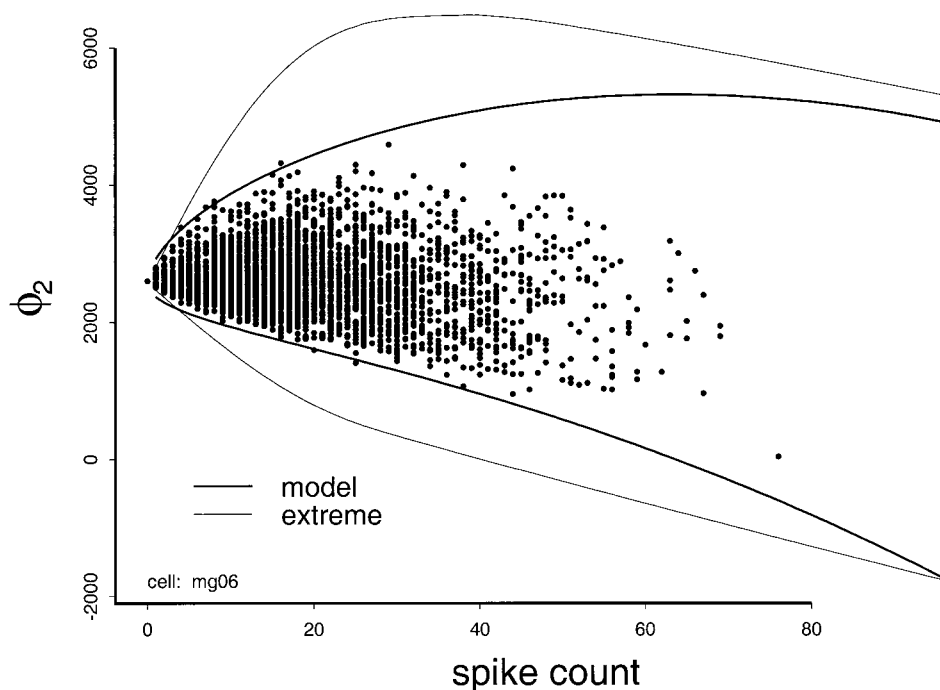


FIG. 5. Spike count- ϕ_2 envelope of a V1 cell. Horizontal axis shows spike count (which is strongly correlated with ϕ_1), the vertical axis ϕ_2 . Dots show spike count and ϕ_2 for individual trials. Thick lines show the spike count- ϕ_2 envelope predicted using the linear relation between $\log(\text{spike count})$ and $\log(\text{variance})$ of ϕ_2 in responses with that spike count, and the quadratic relation between spike count and the mean of ϕ_2 in responses with that spike count, and the assumption that the distribution of ϕ_2 in responses with a given spike count is normal. Thin lines show the spike count- ϕ_2 envelope calculated by taking the largest (or smallest) possible projection onto the ϕ_2 waveform of responses with a given number of spikes. More extreme estimate does not take into account when spikes are most likely to occur or even that it is unlikely that many spikes will occur at nearly the same time. Spike count is used here instead of ϕ_1 for ease of comparison with the extreme bounds.

mean(ϕ_2) from the model. Using these regressions, we predicted the variance of ϕ_1 and ϕ_2 distributions elicited by different stimuli from the means of the ϕ_1 distributions.

To check that our regression estimates were robust, we divided the data for each neuron into two sets. The regressions for the two halves were statistically indistinguishable ($P > 0.01$) for all 54 neurons. The fact that different subsets of the data result in indistinguishable regression lines shows that the model can generalize, and supports using the model to predict the structure of responses to stimuli other than those presented in our experiments.

Distributions of principal component coefficients

Estimating the transmitted information between the visual stimuli and the neuronal responses requires estimating the conditional response distributions $p(\phi_1|s)$ and $p(\phi_2|s)$. We assumed that the ϕ_1 and ϕ_2 distributions are separable (except for the relation between ϕ_1 and the range of ϕ_2) and examined normal, lognormal, and gamma distributions, each truncated at the bounds of the response space, as models for the ϕ_1 and ϕ_2 distributions.

Using the observed mean and the variance estimated from the regression relation, the normal distribution (truncated at the boundary of the response space) was an acceptable fit for 79% of the distributions of ϕ_1 elicited by individual stimuli, 83% of the distributions of ϕ_2 elicited by individual stimuli, and 79% of the distributions of ϕ_2 given ϕ_1 regardless of stimulus (Kolmogorov-Smirnov test, $P < 0.05$). The gamma distribution was acceptable in almost exactly the same cases as the normal distribution, but the lognormal distribution was rejected much more frequently. We chose to use the normal distribution. Because the gamma distribution fit nearly as well as the normal, the information calculations presented in the following text were repeated using the gamma distribution for several cells; the results were indistinguishable from those obtained using the normal distribution.

Transmitted information

In this study, our goal was to estimate channel capacity, not transmitted information (which has been estimated for these data in the past) (see Eskandar et al. 1992; Heller et al. 1995; Kjaer et al. 1997; Richmond et al. 1990). Transmitted information was calculated as a test of our model: estimates based on the model can be compared with estimates obtained using a previously validated neural network method (Golomb et al. 1997; Kjaer et al. 1994). If the assumptions in our model are reasonable, the two methods should give similar results. This has been shown to be the case when spike count is used as the neural code (Gershon et al. 1998); we found that the two methods also give similar results for the (ϕ_1, ϕ_2) code used here. The least-squares line relating the two information estimates had intercept 0.04 and slope indistinguishable from 1. The r^2 value was 0.94. The difference between the two measurements was of the order of magnitude by which Golomb et al. (1997) found that the neural network underestimates transmitted information with limited numbers of samples. Finding that the information measurements using the two methods are similar led us to believe that the assumptions in our model are reasonable and that the model can be used to estimate channel capacity.

Figure 9 shows, for each neuron, the transmitted information using (ϕ_1, ϕ_2) as the neural code plotted against transmitted information using spike count as the neural code. There was no significant difference in the information transmitted using the spike count code by neurons in V1 and IT (V1: 0.35 bits median, interquartile range 0.27–0.55; IT: 0.31 bits median, interquartile range 0.18–0.39; $P > 0.01$ Kruskal-Wallis). Although the increase in transmitted information from the spike count code to the (ϕ_1, ϕ_2) code is significantly larger in V1 than in IT ($P < 0.01$), the difference in information transmitted using the (ϕ_1, ϕ_2) code is still not significant (V1: 0.52 bits median, interquartile range 0.40–0.78; IT: 0.37 bits median, interquartile range 0.22–0.55; $P > 0.01$). This represents an

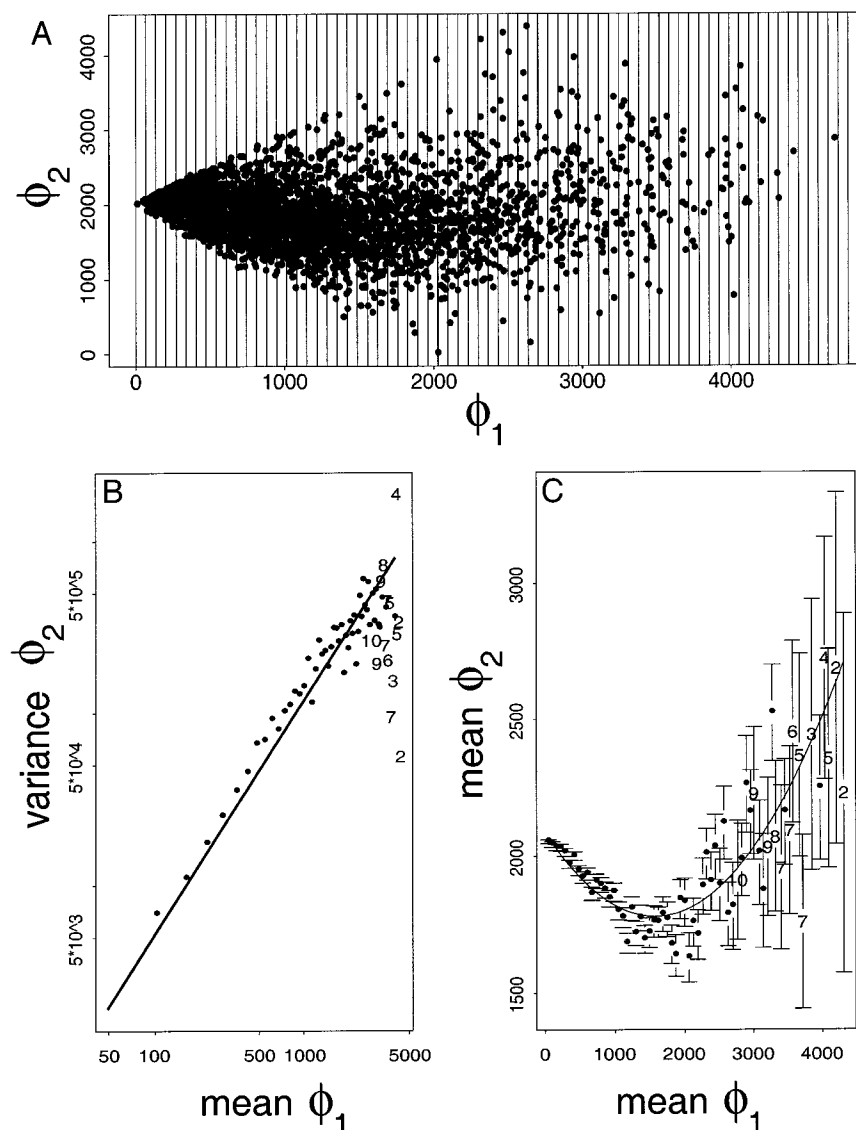


FIG. 6. Predictors of the mean and variance of ϕ_2 for a given range of ϕ_1 . A: dividing the range of ϕ_1 into "slices". Top left scatterplot from Fig. 4 is reproduced with the boundaries of the ϕ_1 slices superimposed. \bullet , (ϕ_1, ϕ_2) for a single response. B: $\log(\text{variance})$ of ϕ_2 values in ϕ_1 slices as a linear function of $\log(\text{mean } \phi_1)$. Horizontal axis shows the mean ϕ_1 value in each slice. Vertical axis shows the variance of ϕ_2 in each slice. \bullet , values in a single ϕ_1 slice. C: mean of ϕ_2 in ϕ_1 slices as a quadratic function of ϕ_1 . Horizontal and vertical axes show the mean values of ϕ_1 and ϕ_2 respectively. \bullet , values in a single ϕ_1 slice. In both B and C, \bullet represent points based on ≥ 11 responses. Points based on fewer responses are represented by the number of responses. Regression lines were calculated using only means and variances based on ≥ 11 responses. Note that the axes in A and C are linear, whereas the axes in B are logarithmic.

increase in transmitted information of 55% (median; interquartile range 22–74%) for neurons in V1, and 19% (median; interquartile range 9–41%) for neurons in IT. Transmitted information using ϕ_1 (which is correlated strongly with spike count) as the code was 12% (median; interquartile range 3–29%) greater than transmitted information using spike count in V1 neurons and 5% (median; interquartile range –1–11%) greater in IT.

To check that we did not lose stimulus-related information by smearing spike arrival times with too broad a convolution kernel, we repeated the information calculations with responses smoothed using a Gaussian kernel with a standard deviation of 1 ms rather than 5 ms. No additional information was found.

Channel capacity

Figure 10 shows, for each neuron, the channel capacity using (ϕ_1, ϕ_2) as the neural code plotted against channel capacity using spike count as the neural code. There was a small but significant difference between the channel capacities using

spike count code of neurons in V1 and IT (V1: 1.1 bits median; interquartile range 0.91–1.30; IT: 1.4 bits median; interquartile range 1.2–1.5; $P < 0.01$ Kruskal-Wallis). There was no significant difference in channel capacity using the (ϕ_1, ϕ_2) code (V1: 2.0 bits median; interquartile range 1.8–2.3; IT: 2.2 bits median; interquartile range 1.8–2.5). This represents an increase in channel capacity of 84% (median; interquartile range 62–124%) for neurons in V1 and an increase of 52% (median; interquartile range 32–95%) for neurons in IT. This increase in channel capacity is a result of temporal modulation and is larger than estimated using only the observed responses (transmitted information). Channel capacity using ϕ_1 (which was correlated strongly with spike count) as the code differed from channel capacity using spike count as the code by 7% (median; interquartile range –7–22%).

We performed several analyses to verify that our estimates of channel capacity are robust with respect to small changes in the response space boundaries. As for spike count code (Gershon et al. 1998), channel capacity depends on the range of responses the cell is capable of emitting in response to a stimulus. If we underestimate a neuron's dynamic range, we

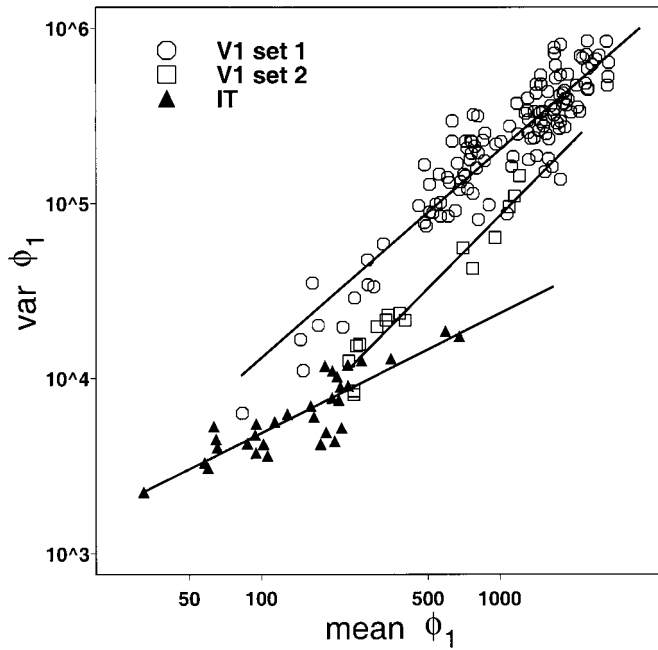


FIG. 7. Log(mean) vs. log(variance) of the distributions of ϕ_1 elicited by different stimuli. Horizontal and vertical axes show (on a logarithmic scale) the mean and variance of ϕ_1 respectively. Each point represents the responses to a single stimulus. There were 128 stimuli for the V1 set 1 neuron (\circ), 16 stimuli for the V1 set 2 neuron (\square), and 32 stimuli for the IT cell (\blacktriangle). Least-squares regression line for each data set is shown. This example shows the cell with the median slope from each data set.

will underestimate its channel capacity. Here we estimated the neuron's dynamic range based on the responses observed. It is possible that we have not used stimuli that elicit the highest possible firing rates (and so ϕ_1) from these neurons. Nonetheless the peak firing rates we saw in these V1 and IT neurons are similar to those reported by others using a wide variety of stimuli, including natural stimuli (Baddeley et al. 1997; Perrett et al. 1984; Rolls 1984; Rolls et al. 1982; Tolhurst et al. 1981, 1983; Vogel et al. 1989), so we believe that our estimate of the dynamic range is reasonable. For several neurons, we examined the effect of allowing part of the distribution of ϕ_1 to fall outside the observed dynamic range. When we allowed as little as 0.5% or as much as 5% of the distribution of ϕ_1 to fall outside the observed range, estimated channel capacity changed by $<4\%$. Similarly, widening the bounds on ϕ_2 for given ϕ_1 by 5% increased the channel capacity by $<3\%$. If new evidence were to show that the proper range for either ϕ_1 or ϕ_2 is larger than we have estimated here, channel capacity could be recalculated using these methods.

Distribution of responses achieving channel capacity

We estimated channel capacity by finding the distribution (in 2 dimensions) of mean (ϕ_1 , ϕ_2) responses that allows the cell to transmit the maximum possible information using a code based on ϕ_1 and ϕ_2 .

Figure 11A shows an example of such a distribution. The horizontal and vertical axes show means of ϕ_1 and ϕ_2 , respectively. Shades of gray indicate how frequently each mean is presented to achieve channel capacity. Because some of these distributions are quite broad, the distribution of observed responses arising from this distribution of mean responses is

diffuse (not shown). The projection of the (ϕ_1 , ϕ_2) distribution onto ϕ_1 , that is, the distribution of mean ϕ_1 implied by the two-dimensional distribution, is shown as a histogram immediately below. Figure 11B shows the distribution of mean ϕ_1 values that achieves channel capacity using ϕ_1 alone as the neural code. In both histograms the horizontal axis shows mean ϕ_1 values, and the vertical axis shows the frequency with which distributions with the appropriate mean value are presented to achieve channel capacity. The projection of the optimal two-dimensional distribution onto the ϕ_1 axis is less concentrated than the optimal one-dimensional distribution.

Role of further principal components

Throughout this study we limited our analysis to two principal components. The reason was practical: using a three-component code increases the computational burden beyond the resources currently available to us. We can, however, address the issue of whether the use of more principal components in the response representation can be expected to lead to substantially higher estimates of information or channel capacity. Successive principal components, by definition, account for successively smaller portions of the response variance, that

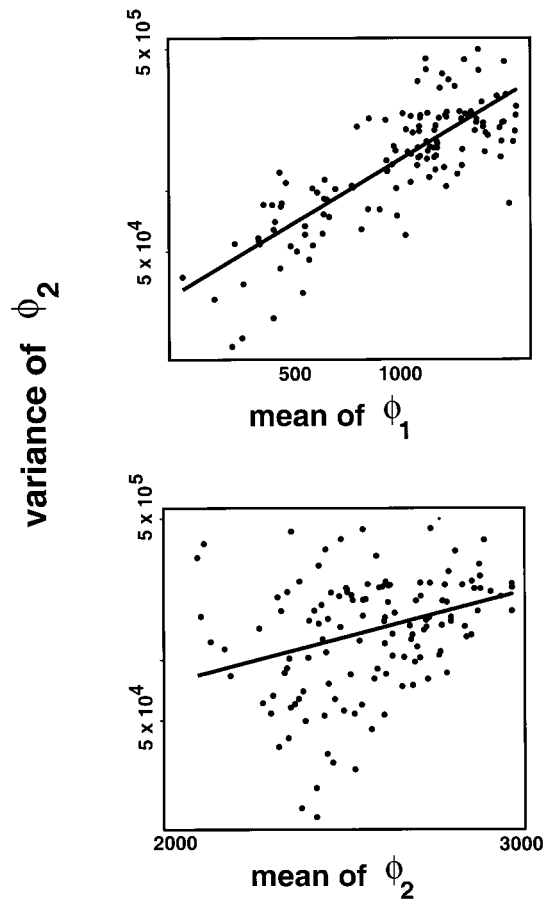


FIG. 8. Variance of ϕ_2 is better predicted by mean of ϕ_1 than by mean of ϕ_2 . *Top*: variance of ϕ_2 by stimulus plotted against mean ϕ_1 by stimulus. Horizontal axis shows mean ϕ_1 and the vertical axis shows variance ϕ_2 . *Bottom*: variance of ϕ_2 by stimulus plotted against mean ϕ_2 by stimulus. Horizontal axis shows mean ϕ_2 and the vertical axis shows variance ϕ_2 . Both axes are logarithmic. \bullet , responses elicited by a single stimulus. Regression lines are superimposed.

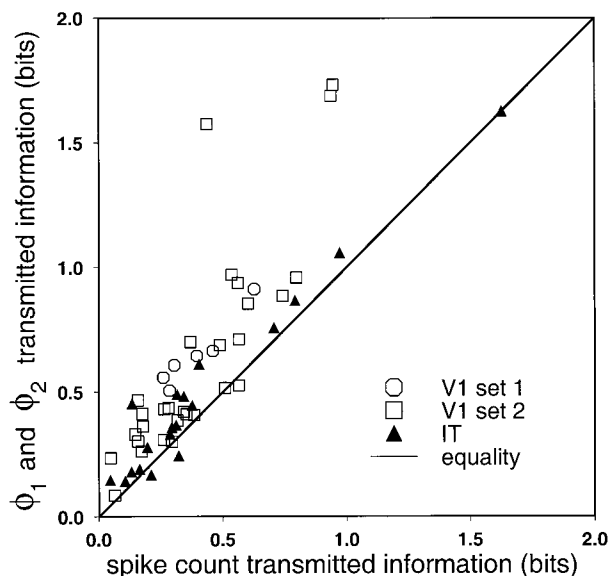


FIG. 9. Transmitted information using spike count or two principal components. Horizontal axis shows transmitted information based on spike count. Vertical axis shows transmitted information based on a code using both ϕ_1 and ϕ_2 . —, equality.

is, their range decreases. This decrease is rapid in our data. Therefore their information content (and so their contribution to channel capacity) must decline unless the noise associated with each principal component also decreases with the range. Figure 12 shows that this decrease does not happen in our data. Each column shows the distribution (across 54 cells) of the signal-to-noise ratio (the variance of mean responses to stimuli divided by the median variance of responses to single stimuli) for spike count or one of the first 10 principal components. A signal-to-noise ratio less than one means that the variability of responses to a given stimulus is greater than the variability of mean responses to different stimuli, so the responses distinguish only poorly among the stimuli that elicited them. Thus the third principal component will contribute much less channel capacity than the first and second principal components, and the fourth and higher principal components are expected to contribute insignificantly if at all. We verified that, as expected, the fifth through tenth principal components contribute no information not redundant with information in the first principal component. Thus our code using only ϕ_1 and ϕ_2 should carry a very large proportion of the information available in the responses.

DISCUSSION

Here we have constructed a model of neuronal responses, based on principal component analysis, that includes both response intensity (firing rate) and a low-precision (± 10 ms) representation of temporal modulation. This temporal precision and this code have been shown to carry a very large portion of the information useful for the identification of statically presented two-dimensional stimuli (Heller et al. 1995; McClurkin et al. 1991; Optican and Richmond 1987; Richmond and Optican 1990; Tovee et al. 1993). Temporal modulation at finer time scales can be used to signal rapid changes in the stimulus (Buracas et al. 1998), but thus far has not been proved useful for stimulus identification.

Transmitted information, which measures how well a set of responses distinguishes among the stimuli actually presented in an experiment, is sensitive to exactly which stimuli were presented and how often. Channel capacity quantifies how useful a neuron with a particular repertoire of responses could be for stimulus identification under the best possible circumstances: increased channel capacity means better stimulus identification. When the repertoire of responses is based on a particular neuron's responses, the calculated channel capacity is an estimate of the neuron's channel capacity. The accuracy of the estimate depends, of course, on how well the model generalizes to responses to stimuli not presented in the experiment. To the extent it was possible to test here, our model generalized well.

Using transmitted information, it has been estimated for some of the data used in this study (Heller et al. 1995) and in other similar studies (Tovee et al. 1993; Victor and Purpura 1996) that temporal modulation adds 0.1–0.2 bits to the information carried by spike count alone. Channel capacity based on the two-principal-component code used here is greater than channel capacity based on the spike count code by 0.8–0.9 bits. This is nearly equivalent (at least in V1) to adding a second independent cell using the spike count code. This is much more than estimated using transmitted information, but far less than would be expected if all degrees of freedom provided by the individual spike times carried independent information. This is true despite the fact that for the code including temporal modulation we allowed our model to give the highest estimates of channel capacity consistent with the response envelope and noise structure estimated from the data (see METHODS).

Representing responses with temporal modulation

Spike count can be thought of as measuring spike arrival times extremely coarsely, recording only whether each spike

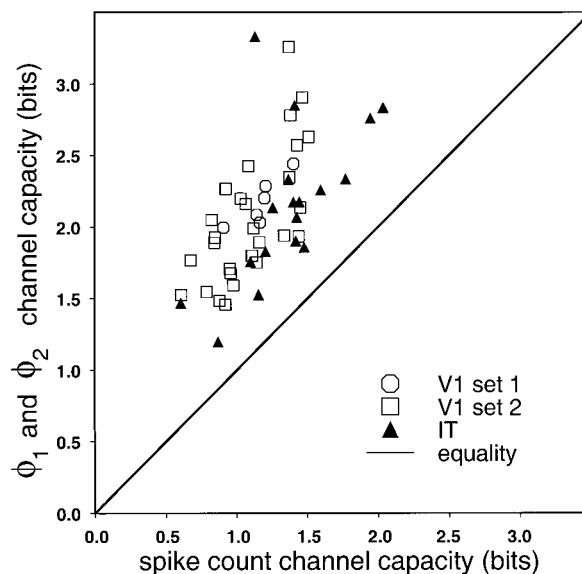


FIG. 10. Channel capacity using spike count or 2 principal components. Horizontal axis shows channel capacity based on spike count alone. Vertical axis shows channel capacity based on a code using ϕ_1 and ϕ_2 . —, equality. Channel capacity using (ϕ_1 , ϕ_2) is always greater than channel capacity using spike count alone.

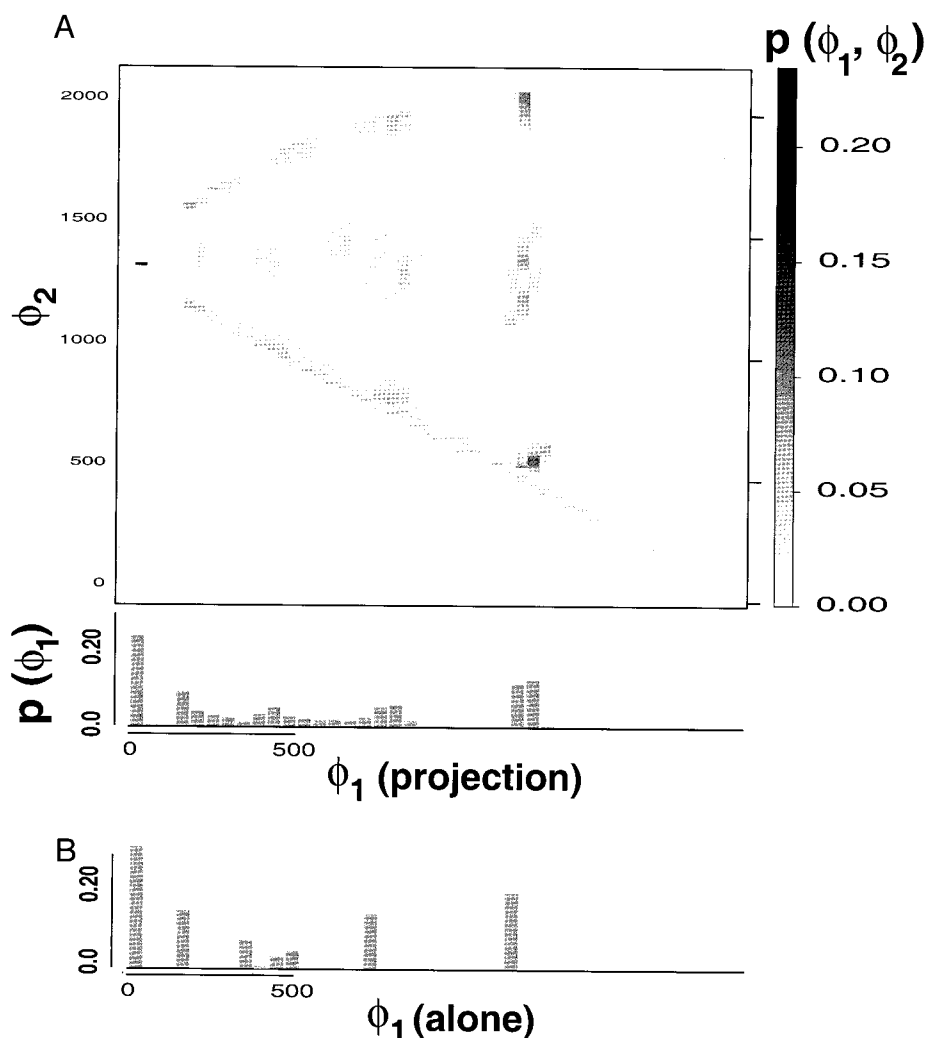


FIG. 11. Distributions of responses that achieve channel capacity. A: distribution of mean (ϕ_1, ϕ_2) pairs that achieves channel capacity based on ϕ_1 and ϕ_2 . Mean values of ϕ_1 and ϕ_2 are plotted on the horizontal and vertical axes; shades of gray show the probability with which distributions with the given means must be presented to achieve channel capacity. Note that the means of the distributions are clustered near opposite edges of the response space. Because some of these distributions are broad, however, the resulting distribution of observed responses is quite diffuse. A histogram of the projection of the distribution of ϕ_1 and ϕ_2 onto ϕ_1 alone is shown below. B: distribution of mean ϕ_1 that achieves channel capacity using ϕ_1 alone as the neural code. Mean values of ϕ_1 are shown on the horizontal axis. Probability with which each mean ϕ_1 must be presented to achieve channel capacity is shown on the vertical axis. Projected distribution shown in A is less concentrated than this distribution based on the ϕ_1 code alone. Two-dimensional distribution loses some ϕ_1 information because the resulting distribution of ϕ_1 is not optimal. This loss of information is offset by a gain in information from ϕ_2 .

arrived during the sample interval or not. The principal component code considered here incorporates temporal modulation on a somewhat less coarse time scale (± 10 ms). Is even more precise representation of the time course of responses from single neurons useful for transmitting information about stimuli?

Spike times originally were measured with 1-ms precision; one way to increase the precision of the representation is to retain this precision when calculating the spike density function rather than smoothing the data. A number of researchers have reported finding information about rapidly changing or moving stimuli encoded using precisely timed spikes. Bair and Koch (1996) found that coherently moving random dot stimuli elicited spikes timed with precisions ranging from ± 2 to ± 15 ms in neurons in area MT of the monkey brain. Buracas et al. (1998) found approximately the same range of precision (± 3 to plus or minus >10 ms) in area MT in response to rapidly moving Gabor gratings. Bair and Koch (1996) did not examine the information content of responses; Buracas et al. (1998) did, but did not report the least temporal precision necessary to encode the information they found (although they indicate that a bin size of 8 ms, corresponding to a precision of ± 4 ms, is adequate).

When information about the identity of a stationary stimulus

has been examined in the responses of monkey cortical neurons, however, less temporal precision has been found. Heller et al. (1995) considered a variety of codes with a broad range of temporal precisions and found that the maximal information was carried by a principal component code cut off at a bandwidth of ~ 25 Hz in V1 and ~ 10 Hz in IT, corresponding to measurement precision of approximately ± 10 ms in V1 and ± 20 ms in IT. Hertz and Richmond (1997) found that the first spike in responses in V1 was placed with a precision of approximately ± 15 ms. Victor and Purpura (1996), using a different approach, found precisions ranging from ± 5 to ± 15 ms in neurons in areas V1 and V2. More recently Oram et al. (1999) confirmed directly that high-precision temporal patterns in neuronal responses from both V1 and LGN are stochastically related to the spike count and (low-bandwidth) poststimulus time histogram and therefore can carry no information beyond that available from the spike count and poststimulus time histogram. In our own data, we found that retaining greater temporal precision in the responses revealed no additional information. We conclude that the information carrying capacity using temporal modulation at the precision used here (± 10 ms) is likely to include almost all of the information carrying capacity available at any time scale for identifying stationary visual stimuli.

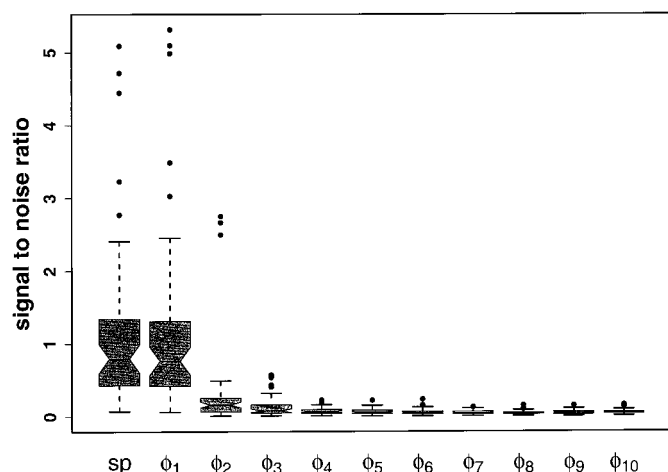


FIG. 12. Fourth and higher principal components of responses are noisy. Boxplot showing the distribution (across 54 cells) of the signal-to-noise ratio for spike count and for the first 10 principal component coefficients. The signal-to-noise ratio is the variance of mean responses to stimuli divided by the median variance of responses to single stimuli. Notches in each box show 95% confidence intervals on the median; the edges of the box show the 25th and 75th percentiles of the distribution, and the ends of the extended lines show the 5th and 95th percentiles. A value below 1 means that the variability of responses to a given stimulus is greater than the variability of mean responses to different stimuli, so the responses distinguish poorly among the stimuli that elicited them. Therefore principal components above the 4th are not expected to carry information.

Neuronal responses also might be represented more precisely using more principal components. Successive principal components have more zero crossings (Fig. 2; see also McClurkin et al. 1991; Richmond and Optican 1987, 1990), so they effectively code higher-frequency fluctuations in the spike density function and so allow greater localization of when spikes occurred. However, successive principal components, by definition, account for successively smaller portions of the response variance. Therefore their information content (and so their contribution to channel capacity) must decline unless the noise associated with each principal component also decreases with the range. We have shown that this does not happen in our data (Fig. 12). A variety of studies confirm that a few principal components carry almost all of the information available in the principal component representation. Richmond and Optican (1990) found that the third principal component increased information transmission by $\sim 5\%$ over that transmitted by the first two principal components for neurons in V1, whereas Tovee et al. (1993) found an 8% increase. McClurkin et al. (1991) found a 7% increase for neurons in the lateral geniculate nucleus. For the data in Heller et al. (1995), the third principal component adds 8% to the information carried by ϕ_1 and ϕ_2 in V1 and $<1\%$ in IT, and the fourth principal component adds $<1\%$ in each area. Therefore we believe that the first two principal components capture most of the temporal coding capability of these neurons.

Principal components are a powerful statistical tool for representing neuronal responses. They identify the features of responses that tend to change most, leading to representations that are efficient for both data compression and information transmission (Campa et al. 1995; Deco and Obradovic 1996; Linsker 1988; Plumbley 1991). However, the fact that information is available in the principal component coefficients does not mean that information is either used or transmitted in

this form. Codes based on principal components are linear codes, and linear codes form a small subset of possible neural codes. The brain may use nonlinear codes that would not be efficiently conveyed using principal components. Latency, for example, which is not concentrated in any single principal component, has been shown to carry information about the luminance and contrast of visual stimuli (Gawne et al. 1996; Mechler et al. 1998; Wiener et al. 1998). Information transmission by spike patterns such as bursts also has been studied (Cattaneo et al. 1981; DeBusk et al. 1997; Reinagel et al. 1999). Nonlinear codes are not necessarily more effective representations of neuronal spike trains than linear codes (Fotheringham and Baddeley 1997), but we cannot rule out that some nonlinear code might carry more information than we have found here.

Statistical model

The response model used here relies on the linear relationship between $\log(\text{mean})$ and $\log(\text{variance})$ of ϕ_1 . This relation is similar to the frequently observed linear relation between $\log(\text{mean})$ and $\log(\text{variance})$ of spike count (Dean 1981; Gershon et al. 1998; Lee et al. 1998; O'Keefe et al. 1997; Tolhurst et al. 1981, 1983; van Kan et al. 1985; Vogel et al. 1989). We identified a similar relationship between the $\log(\text{variance})$ of ϕ_2 and $\log(\text{mean})$ of ϕ_1 , allowing simple modeling of ϕ_2 response variance. ϕ_1 and ϕ_2 were modeled as normal distributions (truncated at boundaries estimated from the data; see METHODS) with the observed mean and variance calculated using the measured regression relation.

The model abstracts away some of the details of the observed responses, and it is reasonable to ask whether the modeled responses are sufficiently like the actual responses to be useful. Estimates of transmitted information using our model were close to those from a previously validated neural network method (Golomb et al. 1997; Heller et al. 1995; Kjaer et al. 1994). This leads us to believe that the model is sufficiently accurate for the information calculations undertaken here. If a better model of the distributions is found, calculations similar to those presented here can be used to recalculate transmitted information and channel capacity.

The response model necessarily is based on the responses actually observed. It is not certain that the model accurately predicts the responses that might be elicited using other stimuli. Several potential problems with generalizing from observed responses can be identified. First, the responses elicited using the Walsh patterns used here might not indicate the full dynamic range of the neurons. We consider this unlikely, because the ranges of spike counts seen in V1 and IT neurons in these experiments were similar to the ranges reported by others using a wide variety of stimuli, including natural stimuli such as hands and faces (Baddeley et al. 1997; Perrett et al. 1984; Rolls 1984; Rolls et al. 1982; Tolhurst et al. 1981, 1983; Vogel et al. 1989). Another potential difficulty is that the logarithms of the mean and variance of responses elicited by other stimuli might not follow a linear relationship or they might follow a linear relationship different from the one measured using these stimuli. Assuming a linear model seems reasonable because a linear relationship between the logarithms of the means and variances of spike count has been observed in many parts of the brain (Dean 1981; Gershon et al.

TABLE 1. *Simplifications used in the response model*

Simplification for Modeling	Evidence	Could it Be Changed?
Responses distributed as modified Gaussians	Based on the data; presented in RESULTS	Given evidence for other distributional forms, could use them
No hidden variables: variances of ϕ_1 and ϕ_2 depend only on the means of ϕ_1 and ϕ_2	Mean-variance relations for spike count, ϕ_1 , and ϕ_2	Difficult; a model with additional stochasticity might be possible
ϕ_1 and ϕ_2 are uncorrelated for each stimulus (except as imposed by response space boundaries)	No consistent correlations seen except at boundaries	Given evidence for correlations, easy to incorporate them
Variance of ϕ_2 depends only on mean of ϕ_1 , not on mean of ϕ_2	Based on the data; presented in RESULTS	Must specify means of ϕ_1 and ϕ_2 , so could easily incorporate influence of mean (ϕ_2)

1998; Lee et al. 1998; O'Keefe et al. 1997; Tolhurst et al. 1981, 1983; van Kan et al. 1985; Vogel et al. 1989). Furthermore we cross-validated the model by dividing the stimuli in half and comparing the resulting mean-variance relations. In all cells, the two relations were indistinguishable, supporting the model's ability to generalize, at least in this limited sense. All of our stimuli were Walsh patterns, and we are unaware of any studies comparing the variability of responses elicited by different kinds of stimuli. If responses elicited by other stimuli are substantially more or less variable than those elicited by the Walsh patterns, our model may not generalize well to responses elicited by other stimuli.

We summarize some of the simplifications made in our model in Table 1.

Spike count, temporal modulation, and channel capacity

Spike count and temporal modulation are linked at the most basic level. When there are no spikes, there can be no temporal modulation. Trains with larger numbers of spikes can form larger numbers of distinct temporal patterns: in the crudest estimate, the number of possible temporal patterns involving k spikes is $n!/[k!(n-k)!]$, where n is the number of distinct time bins. If each pattern could occur with equal probability, the number of temporal patterns observed, and therefore the range of ϕ_2 , would grow rapidly with spike count. If the train becomes crowded—that is, when there are more than half as many spikes as bins—the range decreases again, but this is not an issue in our data.

The range of ϕ_2 does increase with increasing spike count and ϕ_1 (Fig. 4) but more slowly than would be expected if all spike trains with k spikes were equally likely. ϕ_2 is the inner product of a spike density function with Φ_2 , the second principal component of the data. Because Φ_2 tends to have a narrow peak (Richmond and Optican 1987), an extreme positive or negative value of ϕ_2 would indicate that many spikes occur in adjacent bins in a response. Such tight bunching of spikes did not occur in these data. In fact, such bunching is unlikely to occur in neurons with a refractory period. This may give at least a partial understanding of why the observed range of ϕ_2 is smaller than would be expected if arbitrary patterns of spikes were possible (Fig. 5).

Within the constraints estimated from the data, we biased our model to increase estimates of channel capacity due to temporal coding. We did not penalize distributions for probability weight falling outside the response envelope bounds for ϕ_2 (as we did for probability weight falling outside the observed range of ϕ_1); we simply ignored that portion of the

distribution. This allows the widest possible separation between distributions, increasing the estimate of channel capacity. This deliberate overestimation was undertaken to obtain an upper bound on channel capacity. Estimates of transmitted information were unaffected, because the distributions of observed means and variances fell within the response envelope.

Comparison with other studies

Some researchers have reported that rapidly moving or rapidly changing stimuli elicit large numbers of precisely timed spikes in the lateral geniculate nucleus and area MT, transmitting, in some cases, >100 bits/s of information about the changing stimulus (Buracas et al. 1998). However, only much lower precision timing (± 10 ms) and much lower information transmission rates (up to ~ 4 bits/s) have been measured using static stimuli (Heller et al. 1995). Our estimates of channel capacity (~ 2.2 bits in a 330-ms window or ~ 6.6 bits/s), which were designed to err on the side of generosity, fall short of even 10 bits/s (even allowing for a large, and unlikely, contribution from additional principal components). A major difference between these approaches is that in the former, but not the latter, the neuron is essentially entrained by the rapid changes in the stimulus. Theunissen and Miller (1995) have referred to this as temporal coding of a signal, “characterized by a one-to-one correspondence between the time of occurrence of a sensory event and the time of occurrence of the corresponding neural process,” distinguishing it from temporal *encoding* of a signal, in which “information about static or dynamic signals is encoded in some aspect(s) of the temporal pattern of action potentials” without the action potentials being tied to changes in the signal.

The effects of entrainment can be seen even during experiments with static stimuli. Information transmission rates as high as 30 bits/s occur during the onset of responses immediately after presentation of static stimuli (Heller et al. 1995), though these rates drop after ~ 50 ms. Thus the high rate of information transmission is not sustained during typical fixations (of ~ 300 ms) between saccades in normal vision. Nonetheless Richmond et al. (1999) found that in V1 neurons the total amount of information available about the identity of stimuli played one after another in a movie was greater when each frame was presented for 170 ms than when each frame was presented for 136 ms. Elsewhere this information has been found to plateau ~ 150 – 200 ms after stimulus presentation (Gershon et al. 1998). Thus maintaining the highest possible information transmission rate by rapidly changing the stimuli

may not be the best strategy if the goal is to identify objects in the visual field.

Summary

We have constructed a model of responses including both response strength and temporal modulation of firing rate. This model clarifies the implications of assuming that the relations between the means and variances of response features observed during an experiment are representative of the neuron's responses in general. The central feature of the model is that the variances of response features are predictable from the means. This model demonstrates that the variances of both response strength and temporal modulation depend on the mean response strength, with important implications for neural information processing. As we have emphasized, use of the model imposes dependence on how well the model generalizes to responses not observed in our experiments. The results here raise two conflicting possibilities. The first is that the experiments were not designed so as to display the best capabilities of these neurons. The other is that our assumption that all possible responses within the two-dimensional response space actually can arise may not be true. This naturally leads to further experiments to verify under what conditions assumptions made about these features remain valid. A critical question for the future is whether response properties elicited by stimuli of one kind (here, Walsh patterns) are the same in responses elicited by stimuli of other kinds.

APPENDIX

The search for the maximizing set of probabilities is subject to three constraints: the probabilities must be nonnegative; the probabilities must sum to one; and the range of means must be finite. The first two constraints arise from intrinsic properties of probability distributions. If the third constraint is violated, the transmitted information can be infinite and the problem of maximizing transmitted information is ill-posed. The maximization of Eq. 6 is performed numerically. The numerical implementation requires that we discretize the continuous space of mean responses. We denote these discretized probabilities by

$$\bar{p}(\mu_1) = \int_{\mu_1}^{\mu_1 + \Delta\mu_1} d\eta_1 p(\eta_1)$$

$$\bar{p}(\mu_2) = \int_{\mu_2}^{\mu_2 + \Delta\mu_2} d\eta_2 p(\eta_2)$$

The first and second constraints are implemented by requiring that $0 \leq \bar{p}(\mu) \leq 1$ for all μ , and that $\sum_{\mu} \bar{p}(\mu) = 1$, respectively. The third constraint is implemented by penalizing distributions of means that lead to distributions of ϕ_1 that are inconsistent with observed values of ϕ_1 (recall that ϕ_1 and ϕ_2 denote the coefficients with respect to the first and second principal components, respectively). The definition of the conditional distributions (Eq. 2) forces the estimated distribution of responses to fall inside the estimated ϕ_1, ϕ_2 envelope.

Specifically, if $\phi_{1,\min}$ and $\phi_{1,\max}$ are the minimum and maximum permitted ϕ_1 values for a distribution, then we demand that

$$\sum_{\phi_1 > \phi_{1,\max}} C_+(\phi_1 - \phi_{1,\max})p(\phi_1) + \sum_{\phi_1 < \phi_{1,\min}} C_-(\phi_{1,\min} - \phi_1)p(\phi_1) \leq \epsilon \quad (A1)$$

where $p(\phi_1)$ is defined by

$$p(\phi_1) = \int_{M_1, M_2} d\mu_1 d\mu_2 p(\mu_1, \mu_2) p(\phi_1 | \mu_1, \mu_2) \quad (A2)$$

both $C_+(x)$ and $C_-(x)$ are nondecreasing functions of x , and ϵ is small. Eq. A2 ensures that $p(\phi_1)$ falls off rapidly for values of the principal component coefficients outside the permitted range. To implement the optimization procedure, we need to translate this into a constraint on $\bar{p}(\mu)$ because the search for the maximum value of the transmitted information occurs in $\bar{p}(\mu)$ space. Defining the function

$$C(\eta) = \sum_{\phi_1 > \phi_{1,\max}} C_+(\phi_1 - \phi_{1,\max})p(\phi_1 | \eta) + \sum_{\phi_1 < \phi_{1,\min}} C_-(\phi_{1,\min} - \phi_1)p(\phi_1 | \eta) \quad (A3)$$

and combining equations A1 and A2, we arrive at

$$\sum_{\eta} C(\eta) \bar{p}(\eta) \leq \epsilon \quad (A4)$$

which represents our third constraint. In our numerical calculations, we use $\epsilon = 0.01$.

The choice of the functions C_+ and C_- involves a scaling issue not present when calculating channel capacity based on spike count (Gershon et al. 1998). Because the spike count takes on only integer values it has a natural scale that can be used in the penalty functions C_+ and C_- . The principal components, however, have no such natural scale. Dividing all the principal component coefficients by some constant should not change the information or channel capacity, and therefore we must make sure that it does not change the values of C_+ and C_- either. We have chosen to scale C_+ and C_- by the bin size used for ϕ_1 in the numeric calculations.

$$C_+(\phi_1 - \phi_{1,\max}) = \left(\frac{\phi_1 - \phi_{1,\max}}{\text{bin size}} \right)^2$$

$$C_-(\phi_{1,\min} - \phi_1) = \left(\frac{\phi_{1,\min} - \phi_1}{\text{bin size}} \right)^2$$

This means that each of the penalty functions grows roughly as if the bins represented spike count. Note that this means that using a larger number of more narrow bins to cover a given range allows less of the probability mass to fall outside the observed range than using a smaller number of wider bins. However, it does guarantee (as for spike count) that ϵ represents the largest portion of the (unweighted) probability mass that can possibly fall outside the estimated range (based on the observed values as described above). This scaling problem arises only when we want to penalize responses in a way that grows as the responses fall further out of the observed range. If the penalty weight is equal for all probability mass falling outside the estimated range, the scale does not matter.

To find the channel capacity, we maximize transmitted information (Eq. 6) under the constraints discussed in the preceding text. Any standard minimization algorithm can be used. We used a sequential quadratic programming method (Lawrence et al. 1997). We first found the distribution that achieved channel capacity for ϕ_1 alone. We uniformly smeared this distribution over the ϕ_2 dimension for each value of ϕ_1 to create a two-dimensional distribution, which was used as the starting point of the minimization. Because the search space is closed and convex, and transmitted information is a convex function of the probabilities, we are guaranteed a single global minimum (Cover and Thomas 1991). In two neurons, we nonetheless verified that the minimization converged to the same solution from several starting points. We also checked in two cells that increasing the resolution of the discretization of ϕ_1 and ϕ_2 did not significantly change our results.

The authors thank Drs. Peter Latham (University of California, Los Angeles), Zheng Liu (National Institute of Mental Health), and Mike Oram (School of Psychology, University of St. Andrews) for helpful comments on the manuscript.

Address for reprint requests: B. J. Richmond, Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Building 49, Room 1B-80, Bethesda, MD 20892-4415.

Received 20 April 1999; accepted in final form 28 July 1999.

REFERENCES

- AHMED, N. AND RAO, K. R. *Orthogonal Transforms for Digital Signal Processing*. Berlin: Springer-Verlag, 1975.
- BADDELEY, R., ABBOTT, L. F., BOOTH, M.C.A., SENGPTEL, F., FREEMAN, T., WAKEMAN, E. A., AND ROLLS, E. T. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B Biol. Sci.* 264: 1775–1783, 1997.
- BAIR, W. AND KOCH, C. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput.* 8: 1185–1202, 1996.
- BURACAS, G., ZADOR, A., DEWESE, M., AND ALBRIGHT, T. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* 20: 959–969, 1998.
- CAMPA, A., DEL GIUDICE, P., PARGA, N., AND NADAL, J.-P. Maximization of mutual information in a linear noisy network: a detailed study. *Network* 6: 449–468, 1995.
- CATTANEO, A., MAFFEI, L., AND MORRONE, C. Patterns in the discharge of simple and complex visual cortical cells. *Proc. R. Soc. Lond. B Biol. Sci.* 212: 279–297, 1981.
- COVER, T. M. AND THOMAS, J. A. *Elements of Information Theory*. New York: Wiley, 1991.
- DEAN, A. F. The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.* 44: 437–440, 1981.
- DEBUSK, B. C., DEBRUYN, E. J., SNIDER, R. K., KABARA, J. F., AND BONDS, A. B. Stimulus-dependent modulation of spike burst length in cat striate cortical cells. *J. Neurophysiol.* 78: 199–213, 1997.
- DECO, G. AND OBRADOVIC, D. *An Information-Theoretic Approach to Neural Computing*. Berlin: Springer, 1996.
- ESKANDAR, E., RICHMOND, B. J., AND OPTICAN, L. M. Role of Inferior Temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. *J. Neurophysiol.* 68: 1277–1295, 1992.
- FOTHERINGHAME, D. AND BADDELEY, R. Nonlinear principal components analysis of neuronal spike trains. *Biol. Cybern.* 77: 282–288, 1997.
- GAWNE, T. J., KJAER, T. W., AND RICHMOND, B. J. Latency: another potential code for feature binding in striate cortex. *J. Neurophysiol.* 76: 1356–1360, 1996.
- GERSHON, E. D., WIENER, M. C., LATHAM, P. E., AND RICHMOND, B. J. Coding strategies in monkey V1 and inferior temporal cortices. *J. Neurophysiol.* 79: 1135–1144, 1998.
- GOLOMB, D., HERTZ, J., PANZERI, S., TREVES, A., AND RICHMOND, B. J. How well can we estimate the information carried in neuronal responses from limited samples? *Neural Comput.* 9: 649–665, 1997.
- HELLER, J., HERTZ, J. A., KJAER, T. W., AND RICHMOND, B. J. Information flow and temporal coding in primate pattern vision. *J. Comput. Neurosci.* 2: 175–193, 1995.
- HERTZ, J. A. AND RICHMOND, B. J. First-spike timing precision of monkey V1 neurons. *Soc. Neurosci. Abstr.* 23: 1266, 1997.
- KENDALL, M. AND STUART, A. *The Advanced Theory of Statistics. Inference and Relationship*. London: Hafner, 1961, vol. 2.
- KJAER, T. W., GAWNE, T. J., HERTZ, J. A., AND RICHMOND, B. J. Insensitivity of V1 complex cells to small shifts in the retinal image of complex patterns. *J. Neurophysiol.* 78: 3187–3197, 1997.
- KJAER, T. W., HERTZ, J. A., AND RICHMOND, B. J. Decoding cortical neuronal signals: network models, information estimation and spatial tuning. *J. Comput. Neurosci.* 1: 109–139, 1994.
- LAWRENCE, C., ZHOU, J. L., AND TITS, A. *User's Guide for CFSQP: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints*. College Park, MD: Institute for Systems Research, University of Maryland, 1997, version 2.5.
- LEE, D., PORT, N. L., KRUSE, W., AND GEORGOPOULOS, A. P. Variability and correlated noise in the discharge of neurons in motor and parietal areas of primate cortex. *J. Neurosci.* 18: 1161–1170, 1998.
- LINSKER, R. Self-organization in a perceptual network. *Computer* 21: 105–117, 1988.
- MCCURKIN, J. W., GAWNE, T. J., RICHMOND, B. J., AND OPTICAN, L. M. Lateral geniculate neurons in behaving primates i. encoding of visual information in the temporal shape of the response. *J. Neurophysiol.* 66: 794–808, 1991.
- MECHLER, F., VICTOR, J. D., PURPURA, K. P., AND SHAPLEY, R. Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli. *J. Neurosci.* 18: 6583–6598, 1998.
- O'KEEFE, L. P., BAIR, W., AND MOVSHON, J. A. Response variability of MT neurons in macaque monkeys. *Soc. Neurosci. Abstr.* 23: 1125, 1997.
- OPTICAN, L. M. AND RICHMOND, B. J. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* 57: 162–178, 1987.
- ORAM, M. W., WIENER, M. C., LESTIENNE, R., AND RICHMOND, B. J. The stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J. Neurophysiol.* 81: 3021–3033, 1999.
- PERRETT, D. I., SMITH, P.A.J., POTTER, D. D., MISTLIN, A. J., MILNER, A. D., AND JEEVES, M. A. Neurones responsive to faces in the temporal cortex: studies of the functional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* 3: 197–208, 1984.
- PLUMBLY, M. *On information theory and unsupervised neural networks*. Cambridge, UK: Cambridge University Engineering Department, 1991. (Technical Report CUED/F-INFENG/TR 78).
- REINAGEL, P., GODWIN, D., SHERMAN, S. M., AND KOCH, C. Encoding of visual information by LGN bursts. *J. Neurophysiol.* 81: 2558–2569, 1999.
- RICHMOND, B. J., HERTZ, J. A., AND GAWNE, T. J. The relation between V1 neuronal responses and eye movement-like stimulus presentations. *Neurocomputing* 26-27: 247–254, 1999.
- RICHMOND, B. J. AND OPTICAN, L. M. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. Neurophysiol.* 57: 147–161, 1987.
- RICHMOND, B. J. AND OPTICAN, L. M. Temporal encoding of two-dimensional patterns by single units in primate visual cortex. II. Information transmission. *J. Neurophysiol.* 64: 370–380, 1990.
- RICHMOND, B. J., OPTICAN, L. M., PODELL, M., AND SPITZER, H. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J. Neurophysiol.* 57: 132–146, 1987.
- RICHMOND, B. J., OPTICAN, L. M., AND SPITZER, H. Temporal encoding of two-dimensional patterns by single units in primate visual cortex. I. Stimulus-response relations. *J. Neurophysiol.* 64: 351–369, 1990.
- ROLLS, E. Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum. Neurobiol.* 3: 209–222, 1984.
- ROLLS, E. T., PERRETT, D. I., CAAN, A. W., AND WILSON, F.A.W. Neuronal responses related to visual recognition. *Brain* 105: 611–646, 1982.
- SANDERSON, A. C. AND KOBLER, B. Sequential interval histogram analysis of non-stationary neuronal spike trains. *Biol. Cybern.* 22: 61–71, 1976.
- SHANNON, C. E. AND WEAVER, W. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.
- SILVERMAN, B. W. *Density Estimation for Statistics and Data Analysis*. Number 26 in Monographs on Statistics and Applied Probability. London: Chapman and Hall, 1986.
- THEUNISSEN, F. AND MILLER, J. P. Temporal encoding in nervous systems: a rigorous definition. *J. Comp. Neurosci.* 2: 149–162, 1995.
- TOLHURST, D. J., MOVSHON, J. A., AND DEAN, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* 23: 775–785, 1983.
- TOLHURST, D. J., MOVSHON, J. A., AND THOMPSON, I. D. The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Exp. Brain Res.* 41: 414–419, 1981.
- TOVEE, M. J. AND ROLLS, E. T. Information encoding in short firing rate epochs by single neurons in primate temporal visual cortex. *Visual Cogn.* 2: 35–58, 1995.
- TOVEE, M. J., ROLLS, E. T., TREVES, A., AND BELLIS, R. P. Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol.* 70: 640–654, 1993.
- VAN KAN, P.L.E., SCOBAY, R. P., AND GABOR, A. J. Response covariance in cat visual cortex. *Exp. Brain Res.* 60: 559–563, 1985.
- VICTOR, J. D. AND PURPURA, K. P. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol.* 76: 1310–1326, 1996.
- VOGEL, R., SPILLEERS, W., AND ORBAN, G. A. The response variability of striate cortical neurons in the behaving monkey. *Exp. Brain Res.* 77: 432–436, 1989.
- WIENER, M. C., ORAM, M. W., AND RICHMOND, B. Response latency is related to stimulus contrast, but not to response strength. *Soc. Neurosci. Abstr.* 24: 1258, 1998.
- WIENER, M. C. AND RICHMOND, B. J. Using response models to study coding strategies in monkey visual cortex. *Biosystems* 48: 279–286, 1998.